



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2013

Dialäkt Äpp: Dialektologie vermitteln - Dialekte ermitteln

Kolly, Marie-José ; Leemann, Adrian

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-105414>

Conference or Workshop Item

Originally published at:

Kolly, Marie-José; Leemann, Adrian (2013). Dialäkt Äpp: Dialektologie vermitteln - Dialekte ermitteln. In: Jahresbericht des Schweizerischen Idiotikons 2013, Zürich, 31 May 2013. Schweizerisches Idiotikon, 21-39.

ALFRED R. SULZER
Beratungen

Schermengasse 10
7208 Malans/GR
Telefon 081 322 14 14
Nateil 079 285 81 63
Fax 081 322 14 25
a.r.sulzer@bluewin.ch

Dr. PETER OTT
Oberviller-Kirchweg 7
6300 Zug

Revisorenbericht

An den Vorstand
des Vereins für das
Schweizerdeutsche Wörterbuch
Zürich

Sehr geehrte Frau Präsidentin
Sehr geehrte Damen und Herren

In Ausübung des uns übertragenen Mandates haben wir die Jahresrechnung 2013 Ihres
Vereins stichprobenweise geprüft.

Dabei haben wir festgestellt, dass

- Bilanz sowie Gewinn- und Verlustrechnung mit der
Buchhaltung übereinstimmen
- die Buchhaltung ordnungsgemäss geführt ist
- die Aktiven und Passiven nachgewiesen und die Einnahmen und
Ausgaben belegt sind.

Aufgrund des Ergebnisses unserer Prüfung beantragen wir, die Ihnen vorgelegte Rechnung zu
genehmigen.

Zürich, 5. März 2014

Mit freundlichen Grüssen

Alfred R. Sulzer
Dr. Peter Ott

Dialäkt Äpp: Dialektologie vermitteln – Dialekte ermitteln

von Marie-José Kolly, Adrian Leemann

1.0 Einleitung

Wir haben eine innovative Methode entwickelt, um einerseits Dialektologie an eine breite Öffentlichkeit zu vermitteln und andererseits Dialektdaten durch *Crowdsourcing* zu ermitteln – wir möchten Ihnen die iOS-Applikation *Dialäkt Äpp* vorstellen (Leemann & Kolly, 2013; Wemakeit, 2012). *Crowdsourcing* verweist auf «das Auslagern von bisher [...] selbst erbrachten Leistungen auf eine große Anzahl von Menschen über das Internet» (Duden online, 2014).

Wie ist diese Äpp entstanden? Kürzlich hörten Herr und Frau Berner in der Skihütte bei einem Teller Gerstensuppe „Sie haben gerade *Suppecheue* für *Suppenkelle* gesagt – stammen Sie aus Bern?“ Beobachtungen wie diese zeigen, dass Sprecher schweizerdeutscher Dialekte ein hohes Bewusstsein für die geographische Variation ihrer Dialekte haben; zum Beispiel wissen sie in der Regel, dass die Vokalisierung von // zu [u] typisch für die Region um Bern ist (der Konsonant // wird dabei als Vokal [u] ausgesprochen, was in der Sprachwissenschaft als //Vokalisierung bekannt ist). Aus solcher anekdotischer Evidenz entstand die Idee, eine Smartphone-Applikation zu entwickeln, die auf dem Bewusstsein und Interesse der Deutschschweizer für sprachliche Variabilität aufbaut – *Dialäkt Äpp*. Die spielerische Äpp bietet für ihre Benutzer einen hohen Anreiz durch ihre zwei Hauptfunktionen: (1) Sie lässt den Benutzer¹ für 16 Wörter seine dialektalen Aussprachevarianten auswählen und erkennt so seinen Dialekt, (2) sie erlaubt es dem Benutzer, seine Aussprache dieser Wörter aufzunehmen, lädt die resultierenden Audio-Dateien auf einen Server und vernetzt sie auf einer interaktiven Karte. So bietet die Äpp eine auditive Reise durch die schweizerdeutsche Dialektlandschaft. Das Datenkorpus, das über das letzte Jahr durch diese Gratis-

¹ Im Folgenden wird für Benutzerinnen und Benutzer sowie Forscherinnen und Forscher jeweils die männliche Form verwendet – Benutzerinnen und Forscherinnen sind dabei selbstverständlich immer mitgemeint.

iOS-Applikation generiert wurde, enthält Variantauswahl-Daten (aus Funktion (1) abgerufen) von über 42 000 Sprechern und Audio-Daten (aus Funktion (2) abgerufen) von über 2600 Sprechern aus der gesamten deutschsprachigen Schweiz. Die App wurde 58 654 mal heruntergeladen (26. Februar 2014; Veröffentlichung der App: 22. März 2013), war über mehrere Tage die am häufigsten heruntergeladene Gratis-App für iOS-Geräte in der Schweiz (App Annie, 2013) und ist in der deutsch- sowie französischsprachigen Schweiz auf ein breites mediales Echo gestossen (cf. <http://www.pholab.uzh.ch/leuterleemann/app.html>).

Weshalb ist diese App ein interessantes Werkzeug für das Sammeln und Analysieren von Sprachdaten? (1) Die Mikrophone von iOS-Smartphones decken einen breiten Frequenzbereich von 50–20 000 Hz ab, was Audio-Aufnahmen hoher Qualität erlaubt (Faber Acoustical, 2009). Die bisherige Forschung hat gezeigt, dass Audio-Aufnahmen eines iPhone der ersten Generation aus dem Jahr 2007 zuverlässige akustische Messungen von Vokalformanten erlauben, die für Sprachanalysen verwendet werden können (de Decker & Nycz, 2011). Vokalformanten sind akustische Energiemaxima, die massgeblich am Klang eines Lautes beteiligt sind. (2) Die positive öffentliche Resonanz zeigt, dass Benutzer an den Funktionen der *Dialäkt App* Gefallen finden, was die Anzahl der Sprecher, die der App ihre Stimme gespendet haben, potenziert hat. Damit werden Resultate von Parent & Eskenazi (2011) bestätigt: Spielerischer, ansprechender Inhalt führt beim *Crowdsourcing* von Daten zu einer höheren Anzahl Teilnehmer. (3) Forschung zu *Crowdsourcing* zeigt, dass die Angaben der meisten Benutzer verlässlich sind und dass die Menge an Daten einzelne Mängel kompensiert: Die Redundanz der Daten kommt für ihre Qualität auf (Munro, Bethard, Kuperman, Lai, Melnick et al., 2010; Juska-Bacher, Biemann, Quasthoff, 2012). In den durch die *Dialäkt App* gewonnenen *Crowdsourcing*-Daten liegt somit erhebliches wissenschaftliches Potenzial. Zum Beispiel erlaubt das *Dialäkt App*-Datenkorpus flächendeckende Untersuchungen von Sprachwandel in der schweizerdeutschen Dialektlandschaft, da unser Algorithmus zur Dialekterkennung auf historischen Karten des Sprachatlas der Deutschen Schweiz (SDS, 1962–2003) basiert. Durch ihre Hauptfunktionen sammelt die *Dialäkt App* Aussprachedaten in denselben Ortschaften, die in den SDS-Karten verwendet wurden, was einen direkten Vergleich mit historischen Daten erlaubt.

Der vorliegende Artikel ist wie folgt gegliedert: Abschnitt 2 betrachtet die *Dialäkt App* mit Fokus auf den Benutzer, indem die Funktionen der App sowie die Methoden für deren Implementierung vorgestellt werden. Abschnitt 3 betrachtet die *Dialäkt App* mit Fokus auf den Forscher, indem das wissenschaftliche Potenzial des *Dialäkt App*-Korpus diskutiert wird. Abgerundet wird dieser Abschnitt durch eine Präsentation erster Forschungsergebnisse.

2.0 *Dialäkt App* für den Benutzer: Funktionen zur Vermittlung von Dialektologie und Methoden zu deren Implementierung

Die *Dialäkt App* bietet folgende vier Funktionen:

- Automatische Dialekterkennung (Name der Funktion in der App: *Verortung*)
- Ortschaften und ihre Aussprachen – Aussprachen und ihre Ortschaften (*Wie? Wo?*)
- Audio-Aufnahmen (*Aufnahme*)
- Text zum Wort der Woche (*Wort der Woche*)

2.1 Automatische Dialekterkennung

«Welchen Dialekt sprichst du?» – eine häufig gestellte Frage an Apéros und anderen gesellschaftlichen Anlässen in der Deutschschweiz, auf welche die *Dialäkt App* eine Antwort bietet, die vom Benutzer evaluiert werden kann.

Funktion: Für 16 Wörter fordert die *Dialäkt App* ihren Benutzer auf, seine jeweilige dialektale Aussprachevariante – durch Tippen auf den Smartphone-Bildschirm – aus einer Liste auszuwählen. Da die Deutschschweizer Dialekte keine standardisierte Orthographie kennen, wurden Varianten möglichst aussprachegetreu aufnotiert. In Fällen, wo Varianten durch Buchstabierung nicht unterschieden werden können, werden sie anhand phonetischer Transkriptionen differenziert (e. g. *Donschtig* [ɔ], *Donschtig* [o]; cf. Abbildung 1, oben links). Da Vertrautheit mit phonetischer Transkription nicht vorausgesetzt werden kann, bietet die App zusätzlich vorgefertigte Aufnahmen aller Varianten zur auditiven Unterscheidung phonetischer Details. Phonetische Transkriptionen sowie vorgefertigte Aufnahmen erlauben es uns, eine hohe Anzahl Varianten für jede Variable einzubinden, was die Treffsicherheit der Dialekterkennung verfeinern sollte. Jede Variable mit ihren Varianten wird auf einem separaten Bildschirm präsentiert, der im Titel das standarddeutsche Wort enthält und im Untertitel die relevante Variable (e. g. *Erster*

Vokal). Die meisten der 16 Wörter beinhalten ein phonetisches Dialektmerkmal, es wurden aber auch eine lexikalische und eine morphologische Variable eingebunden. Unter dem Untertitel werden die dialektalen Aussprachevarianten aufgelistet, wobei die relevante Variable unterstrichen ist (cf. Abbildung 1, oben links). Tabelle 1 präsentiert die 16 Variablen.

Tabelle 1: Variablen zur Dialekterkennung (mhd. = mittelhochdeutsch), Beispiele, Anzahl und Typen von schweizerdeutschen Varianten (phon. = phonetisch, lex. = lexikalisch, morph. = morphologisch)

Standard-deutsches Wort	Variable	Beispiele für Varianten	N	Typ
<i>Abend</i>	mhd. <u>-ent</u> , <u>-ant</u> , <u>-int</u> , <u>-unt</u>	<i>Aabe</i> , <i>Aabig</i>	13	phon.
<i>Abend</i>	mhd. <u>â</u>	<i>Aabe</i> [a], <i>Aabe</i> [a]	8	phon.
<i>Apfelüberrest</i>	Lexem	<i>Bütschgi</i> , <i>Gröibsch</i>	39	lex.
<i>Augen</i>	mhd. <u>ou</u>	<i>Aige</i> , <i>Ouge</i>	11	phon.
<i>Bett</i>	mhd. <u>e</u>	<i>Bett</i> [e], <i>Bett</i> [ε]	2	phon.
<i>Donnerstag</i>	mhd. <u>o</u>	<i>Donschtig</i> [o], <i>Dunschtig</i> [u]	8	phon.
<i>Kelle</i>	mhd. <u>ll</u>	<i>Chälle</i> , <i>Chäue</i>	5	phon.
<i>Kind</i>	mhd. <u>k</u> , <u>ch</u>	<i>Chind</i> [x], <i>Schind</i> [j]	4	phon.
<i>Tanne</i>	mhd. <u>n</u> , <u>nn</u>	<i>Tane</i> , <i>Tanne</i>	2	phon.
<i>fragen</i>	mhd. <u>â</u> , <u>æ</u>	<i>fraage</i> , <i>fräage</i>	10	phon.
<i>heben</i>	mhd. <u>u</u>	<i>lupfe</i> [u], [v], <i>lupfe</i> [y], [y]	3	phon.
<i>hinauf</i>	mhd. <u>hin-uf</u>	<i>embruf</i> / <i>embrüf</i> , <i>uehi</i>	31	morph.
<i>schneien</i>	mhd. <u>i</u>	<i>schneie</i> , <i>schnije</i>	7	phon.
<i>spät</i>	mhd. <u>â</u>	<i>schpaat</i> [v], <i>schpeät</i> [ea]	12	phon.
<i>tief</i>	mhd. <u>iü</u>	<i>täif</i> , <i>tüüf</i>	7	phon.
<i>trinken</i>	mhd. <u>i</u> , <u>-nk-</u>	<i>treeche</i> , <i>trinkche</i>	10	phon.

Nach der Auswahl von 16 Aussprachevarianten ermittelt der Algorithmus die Liste der fünf Ortschaften, die am besten mit der dialektalen Aussprache des Benutzers übereinstimmen (Abbildung 1, oben rechts), und stellt sie auf einer Karte dar (Abbildung 1, unten links). Benutzer können die Treffsicherheit des ersten Treffers beurteilen (Abbildung 1, oben rechts): Wird dieser als angemessen akzeptiert, so werden die Benutzer gebeten, Alter und Geschlecht anzugeben (Abbildung 1, unten rechts) und damit unsere Forschung zu unterstützen. Wird der

erste Treffer als ungenau oder falsch empfunden, so können Benutzer ihren Dialekt aus einer Liste von Ortschaften auswählen. Die Aussprachevarianten der Benutzer, zusammen mit den Daten zu Alter, Geschlecht und Dialekt, werden anonym auf einem Server abgespeichert.

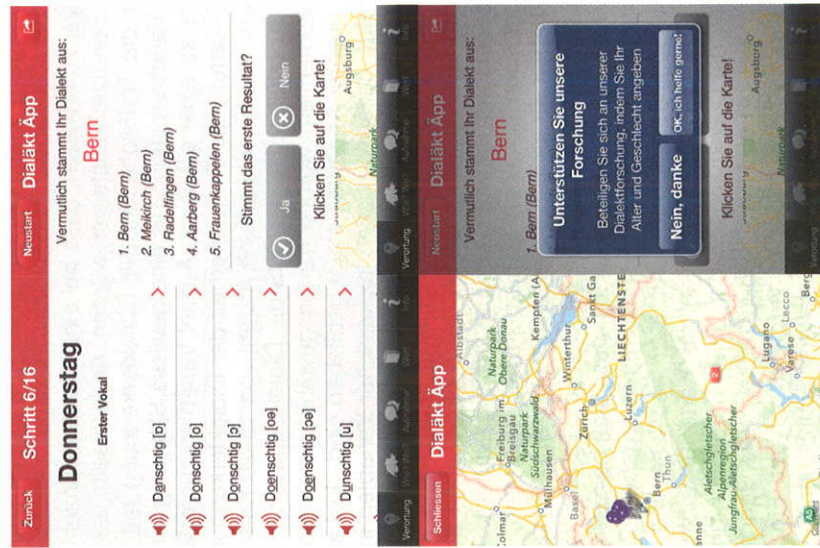


Abbildung 1: Aussprachevarianten für *Donnerstag* (oben links), Dialekterkennung als Liste der fünf besten Treffer (oben rechts) und als Pins auf einer Karte (unten links), Aufforderung an die Benutzer, ihre Daten für Forschungszwecke freizugeben (unten rechts)

Methode: Der Algorithmus zur Dialekterkennung basiert auf 16 Karten des SDS (1962–2003), eines der bedeutendsten Werke der schweizerdeutschen Dialektologie. Der SDS dokumentiert die regionale Variation der Deutschschweizer Dialekte in der ersten Hälfte des 20. Jahrhunderts – die Daten wurden mehrheitlich in den 1940er Jahren gesammelt (wie damals

üblich dienten ältere Sprecher als Gewährspersonen, um ältere dialektale Varianten zu dokumentieren) und decken phonetische, phonologische, morphologische, lexikalische und syntaktische Variation in 566 Deutschschweizer Ortschaften ab. Die Anzahl der für die App ausgewählten Karten, 16, unterliegt einer Abwägung zwischen Benutzerfreundlichkeit und Treffsicherheit der Dialekterkennung. Hinsichtlich der Benutzerfreundlichkeit war zu erwarten, dass bei einer höheren Anzahl Variablen viele Benutzer aussteigen würden. Andererseits nahmen wir an, dass die Treffsicherheit der Dialekterkennung mit der Anzahl der Variablen steigt, bis eine gewisse Sättigung erreicht ist, d. h. dass ab einer gewissen Anzahl Variablen zusätzliche Daten (SDS-Karten) die Treffsicherheit der Dialekterkennung nicht verbessern. Intuitiv schien eine Anzahl von 15–20 Karten diese Kriterien zu erfüllen. Da der SDS die Dialektsituation um 1900–1950 dokumentiert, wurden Karten gewählt, deren Isoglossen gemäss unserer Annahme stabil geblieben sind. Juska-Bacher (2010) zeigt, dass Isoglossen lexikalischer Variablen innerhalb von ca. 50 Jahren starke Veränderungen erfahren haben. Zum Beispiel wurden Varianten wie *Fifalter*, *Summervogel*, *Müllervogel* (wie im SDS dokumentiert) in vielen Ortschaften durch die standarddeutsche Variante *Schmetterling* verdrängt. Phonetische Isoglossen scheinen stabiler zu sein (Christen, 1998) – deshalb haben wir für den Dialekterkennungsalgorithmus hauptsächlich phonetische Variablen gewählt (cf. Tabelle 1). Weiter haben wir die Karten so ausgewählt, dass sich beim Bilden von Schnittmengen maximal kleine Teilmengen ergeben. Zum Beispiel unterscheidet die Karte für *Bett* westliche und östliche Dialekte (Abbildung 2, oben links), die Karte für *schneien* und östliche gegen teilt die Dialektlandschaft in einen nördlichen und einen südlichen Teil (Abbildung 2, oben rechts). Die Schnittmenge dieser beiden Karten bildet im Vergleich zu den Einzelkarten eine verfeinerte Partition. Eine weitere Schnittmenge mit der Karte für *tief* ergibt eine noch höhere Anzahl Teilmengen (Abbildung 2, unten).

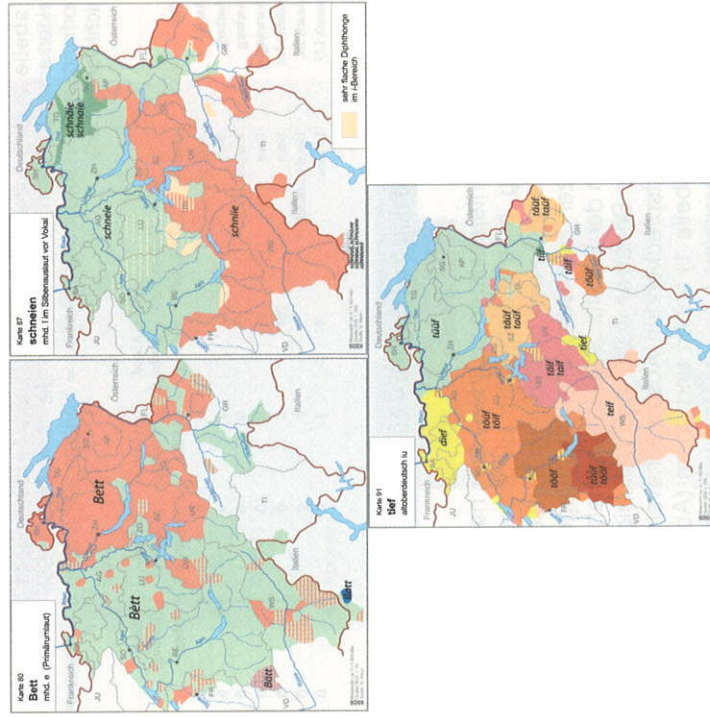


Abbildung 2: Karten für *Bett* (oben links), *schneien* (oben rechts), *tief* (unten), cf. Glaser, Christen, Friedli & Renn (2013)

Die auf den 16 Karten verzeichneten Daten wurden in eine Tabelle konvertiert, welche die Basis für unseren Algorithmus bildet. Die Tabelle enthält eine Zeile pro Ortschaft und eine Spalte für jede Aussprachevariante. Jede Zeile enthält entweder eine 1 (Variante in der Ortschaft dokumentiert) oder eine 0 (Variante in der Ortschaft nicht dokumentiert, cf. Tabelle 2). Der Algorithmus zur Dialekterkennung summiert für jede Zeile alle 1 und gibt die Ortschaft mit der höchsten Summe als besten Treffer aus. Wir haben die im SDS verwendeten Ortschaften übernommen und diejenigen Ortschaften angepasst, welche in der Zwischenzeit fusioniert haben oder umbenannt worden sind, was zu 550 Ortschaften führt (BFS, 2012).

Tabelle 2: Beispielhafte tabellarische Basis für den Algorithmus zur Dialekterkennung. Zeilen dokumentieren die für jede der 550 Ortschaften vorhandenen Aussprachevarianten (1 = Variante vorhanden, 0 = Variante nicht vorhanden)

City	Canton	German word ID / Label	Abend (Endung)		Abend (Endung)	
			Aabig	Aabig	Aabe	Aaben
Aarau	Aargau	34	0	0	0	1
Aarburg	Bern	31	0	0	0	1
Aarburg	Aargau	45	0	0	0	1
Aarwangen	Bern	10	0	0	0	1
Adelboden	Bern	104	0	0	0	1
Aesch BL	Basel-Landsch	7	0	0	0	1
Aesch LU	Luzern	9	1	0	0	0

2.2 Ortschaften und ihre Aussprachen – Aussprachen und ihre Ortschaften

Möglicherweise interessieren sich Benutzer nach der automatischen Erkennung ihres Dialektes für die Menge der Aussprachevarianten, die zu einer bestimmten Ortschaft gehört. So kann der Benutzer zum Beispiel nachschlagen, welche Aussprachevarianten der SDS für „seine“ Ortschaft verzeichnet.

Funktion: Der Benutzer wählt eine beliebige Ortschaft aus und kann die entsprechenden Aussprachevarianten für die 16 Variablen (cf. Tabelle 1) sehen und hören (Abbildung 3, oben links). Weiter interessiert sich der Benutzer vielleicht dafür, wo eine bestimmte Aussprachevariante, zum Beispiel vokalisiertes // in [ʎæwə] *Kelle*, verwendet wird. Nach der Auswahl einer Variable und ihrer Variante (Abbildung 3, oben rechts) zeigt eine interaktive Karte violette Pins für alle Ortschaften, wo die Variante dokumentiert ist (Abbildung 3, unten; cf. SDS, 1962–2003).

Methode: Diese Funktion basiert auf der Tabelle, welche auch die Basis für die Dialekterkennung bildet (Tabelle 2). Entweder extrahiert der Algorithmus für eine bestimmte Ortschaft (Zeile) alle Varianten, die mit 1 codiert sind, oder er extrahiert für eine bestimmte Variante (Spalte) alle Ortschaften, die mit 1 codiert sind, und stellt diese auf einer Karte dar.

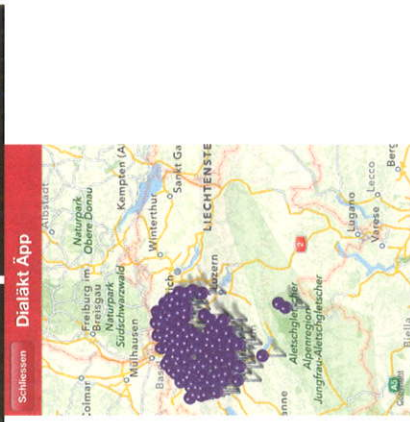


Abbildung 3: Menge der Aussprachevarianten für Basel (oben links), Aussprachevarianten für *Kelle* (oben rechts) und Darstellung der Ortschaften, die eine bestimmte Variante (hier *Chälle*) benutzen (unten)

2.3 Audio-Aufnahmen

Die *Dialäkt App* lädt ihre Benutzer dazu ein, ihre eigene Aussprache der 16 Variablen aufzunehmen, und macht diese auf einer interaktiven Karte sicht- und hörbar. So wird dem Benutzer eine auditive Reise durch die schweizerdeutsche Dialektlandschaft angeboten.

Funktion: Benutzer werden dazu aufgefordert, ihre Stimme in einer maximal ruhigen Umgebung aufzunehmen (Abbildung 4, oben links). Vor der Aufnahme wählen sie ihren Dialekt aus einer Liste von Ortschaften aus und geben ihr Alter und Geschlecht an (Abbildung 4, oben rechts). Danach erscheint je nach der 16 Wörter auf einem separaten Bildschirm und kann

aufgenommen, angehört und, wenn nötig, wieder aufgenommen werden. Weiter können Benutzer ihre eigene dialektale Wiedergabe eines kurzen Dialogs (die ersten beiden Teile des Dialogs *Gespräch am Neujahrstag*, cf. Phonogrammarchiv der Universität Zürich, 2000) aufnehmen. Auf einer interaktiven Karte sind alle vorhandenen Aufnahmen verlinkt (Abbildung 4, unten links). Violette Nadeln stellen Ortschaften mit zeitgenössischen Benutzeraufnahmen dar, grüne Nadeln kennzeichnen Ortschaften, für welche zusätzlich historische Aufnahmen (aus 1943 und 1969) des *Gesprächs am Neujahrstag* verfügbar sind. Benutzer können die Karte heranzoomen, auf eine Nadel, die zu einer bestimmten Ortschaft gehört, tippen, eine Variable auswählen und alle von anderen Benutzern aufgenommenen Varianten für diese spezifische Ortschaft anhören (Abbildung 4, unten rechts).

Methode: Die Audio-Aufnahmen und Metadaten der Benutzer werden anonym auf einen Server geladen und auf einer interaktiven Karte dargestellt. 24 historische Aufnahmen des *Gesprächs am Neujahrstag* erscheinen auf der selben Karte (Abbildung 4, unten links).

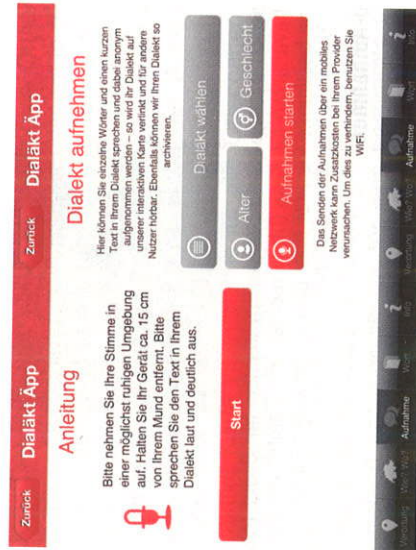


Abbildung 4: Anleitung für die Aufnahme (oben links), Aufforderung an die Benutzer, Metadaten anzugeben (oben rechts), Ortschaften mit zeitgenössischen (violette Nadeln) und historischen (grüne Nadeln) Aufnahmen (unten links), Audio-Aufnahmen von *Donnerstag* für Zürich (unten rechts)

2.4 Text zum Wort der Woche

Viele Smartphone-Applikationen werden nur einmal verwendet und danach vom mobilen Gerät gelöscht (25 %; cf. MobiThin-king, 2013). Wir haben eine Funktion in die Dialekt App eingebaut, welche dieser Tatsache hoffentlich entgegenwirkt – eine Funktion, welche die Benutzer mit wechselnden App-Inhalten anregt. Dies ist attraktiv, da es den Benutzer zur Wiederverwendung der App auffordert. So wird eine längerfristige Nutzung der *Dialekt App* interessant.

Funktion: Jeden Montag erscheint ein kurzer Text, der spezifisch für Nicht-Linguisten kreiert wurde, in der Rubrik *Wort der Woche* (Abbildung 5, oben rechts). Dieser kurze Text enthält etymologische, diatopische, kulturgeschichtliche und weitere Informationen zu einem spezifischen schweizerdeutschen Dialektwort (Landolt, 2012–2013; Abbildung 5, oben links). Diese Rubrik ist bei den Benutzern gemäss den regelmässigen Benutzerkommentaren zur *Dialekt App*, die wir per E-mail erhalten, sowie den Rezensionen der App sehr beliebt (Abbildung 5, unten).

Methode: Die Texte für die Rubrik *Wort der Woche* wurden freundlicherweise vom Schweizerischen Idiotikon zur Verfügung gestellt. Behandelte Wörter oder Themen entstehen aus Ideen des Autors (Landolt, 2012–2013), aus Anfragen an das Schweizerische Idiotikon oder aus Aktualitäten (e. g. Jahreszeit, cf. *Wort der Woche* zu *Chilbi* am 19.08.2013). Quellen für diese

Texte sind das Schweizerische Idiotikon, der SDS, weitere Nachforschungen und der Wissensfundus des Autors (Christoph Landolt, Schweizerisches Idiotikon, persönliche Mitteilung).

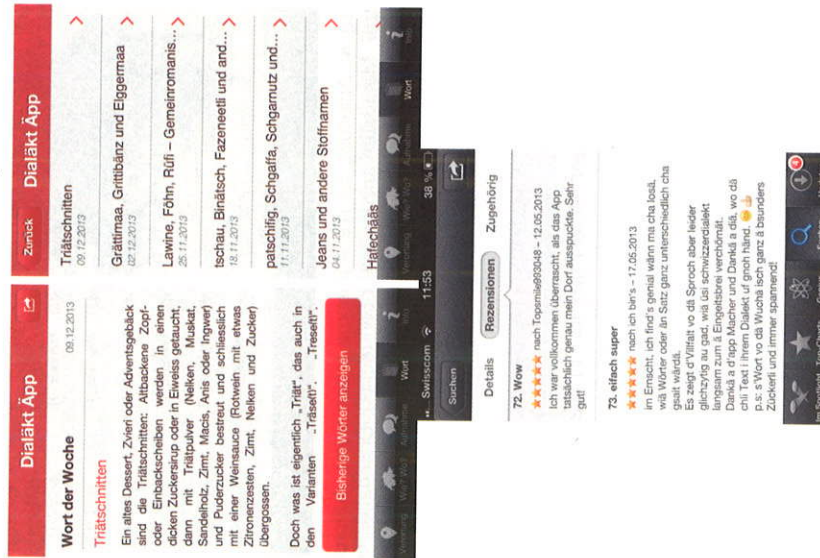


Abbildung 5: Wort der Woche vom 09.-15.12.2013 (oben links), vorherige Wörter (oben rechts) und Kommentar zum Wort der Woche in einer Rezension der Dialäkt App (unten)

3.0 Dialäkt App für den Forscher: Ermittlung von Dialektdaten per Crowdsourcing

Traditionelle Methoden der empirischen linguistischen Forschung basieren ihre Analysen auf eine relativ stark begrenzte Anzahl Sprecher. Sogar neuere *rapid-anonymous*-Studien erreichen typischerweise höchstens ein paar hundert Sprecher (Mather, 2012; Leemann, Kolly, Werlen, Britain & Studer-Joho, im Druck). Um an eine grössere Anzahl Teilnehmer zu gelan-

gen, haben Linguisten vor rund zehn Jahren begonnen, webbasierte Technologien zu nutzen (Siebenhaar, 2003; Vaux & Golder, 2003; Draxler & Steffen, 2005; Schlippe, Ochs & Schultz, 2013) – die grösste Stichprobe, welche so für Deutschschweizer Dialekte erhoben wurde, enthält Daten von über 14 000 Sprechern (Glaser, 2008). In den letzten Jahren haben Forscher aus Fachgebieten wie den kognitiven Wissenschaften (Dufau, Duñabeita, Moret-Tatay, McGonigal, Peeters et al., 2011) oder der Geophysik (Overeem, Robinson, Leijnse, Steeneveld, Horn et al., 2013) erfolgreich Smartphone-Applikationen als Tools verwendet, um per *Crowdsourcing* noch grössere Datenbanken zu kreieren. In der Linguistik ist die Verwendung dieser Technologie neu: de Vries, Davel, Badenhorst, Basson, de Wet et al. (2014) verwenden eine Smartphone-Applikation, um gesprochene Sprache zu elizitieren und damit akustische Modelle für eine automatische Spracherkennung zu trainieren, und das Iwaidja Inyaman Team (2012) hat eine App kreiert, um vom Aussterben bedrohte Sprachen zu dokumentieren.

3.1 Dialäkt-App-Datenkorpus: Beschreibung

Zwei Arten von Daten werden durch die *Dialäkt App* erhoben: Die Dialekterkennungsfunktion sammelt für 16 Wörter pro Benutzer dialektale Aussprachdaten per Variantenauswahl, dazu Angaben zu Alter, Geschlecht und Dialekt der Benutzer (Selbsteinschätzung). Die Aufnahmefunktion erhebt dialektale Audio-Daten von höchstens 16 Wörtern und zwei Dialog-Teilen pro Benutzer sowie dieselben Metadaten. Das Datenkorpus enthält die Variantenauswahl-Daten von 42 681 Benutzern (58 % männlich, 42 % weiblich) und Audio-Aufnahmen von 2633 Benutzern (54 % männlich, 46 % weiblich; Stand 28. Februar 2014). Die meisten Benutzer sind zwischen 18 und 41 Jahre alt und sie sprechen mehrheitlich Dialekte aus den Kantonen Zürich, Bern, Aargau, St. Gallen und Luzern (über 65 %). Die Daten sind flächendeckend verteilt: Variantenauswahl-Daten sind für 547, Audio-Daten für 428 von 550 Ortschaften vorhanden, wobei für die Variantenauswahl-Daten oft über 100 Sprecher pro Ortschaft vertreten sind. Eine vorläufige Beurteilung unseres Dialekterkennungsalgorithmus zeigt, dass der erste vom Algorithmus angegebene Treffer für 64 % der Benutzer im selben Kanton (also annähernd in der selben Dialektregion) liegt wie die vom Benutzer angegebene Ortschaft seines Dialekts.

3.2 *Dialäkt-Äpp*-Datenkorporus: wissenschaftliches Potenzial

Diese Daten können vom Dialektologen dazu verwendet werden, um (areale) Variation zwischen Ortschaften zu untersuchen, und vom forensischen Phonetiker, um Variationsmodelle innerhalb von Ortschaften (d. h. zwischen Sprechern) zu erstellen. Forensisch-phonetische Experten vergleichen typischerweise die Stimmaufnahme eines Täters mit der eines Verdächtigten, um zu ermitteln, wie ähnlich die beiden Stimmen sind.

Mit dem *Dialäkt-Äpp*-Korporus ist der Dialektforscher dazu in der Lage, die areale Verteilung von Varianten zu studieren und entsprechende Karten zu erstellen oder Distanzen zwischen verschiedenen Dialektregionen zu quantifizieren, wie dies für Varietäten des amerikanischen Englischen (Katz, 2013) und des Deutschen (Lameli, 2004; Elspass & Möller, 2003–2014) getan wurde. Die *Dialäkt-Äpp*-Daten erlauben zusätzlich einen Vergleich mit den arealen Verteilungen, die im SDS dokumentiert wurden. So kann die Dialektlandschaft um 1900–1950 mit der Dialektlandschaft um 2013 verglichen werden, womit wir Sprach- und insbesondere Lautwandelphänomene untersuchen können. Das *Dialäkt-Äpp*-Korporus ermöglicht also eine systematische Untersuchung der Stabilität von phonetischen, lexikalischen und morphologischen Isoglossen, basierend auf einer kleinen Auswahl von Variablen und einer grossen Anzahl von Sprechern bei hoher Ortsdichte. Eine erste Analyse der Variantenauswahl-Daten von 42 681 Benutzern zeigt, dass die Dialektvarianten für das Lexem *Apfelüberrest* und für morphologische Varianten von *hinauf* seit der SDS-Erhebung geographischem Wandel stärker ausgesetzt waren – zum Beispiel hat die Zürcher Variante *Bütschgi* für *Apfelüberrest* an Terrain gewonnen – als die dialektalen Varianten für den Vokal in *tief* und die Länge von */n/* in *Tanne*. Für *Apfelüberrest* und *hinauf* benutzen nur 53 % respektive 58 % der Benutzer noch dieselben Varianten, die für ihre Ortschaft im SDS dokumentiert sind. Für die phonetischen Variablen in *tief* und *Tanne* hingegen verwenden 69 % respektive 83 % der Benutzer immer noch die im SDS verzeichnete Aussprache. Erste Resultate stimmen somit mit Christen (1998) und Juska-Bacher (2010) überein: Phonetische Variablen scheinen über die Zeit hinweg stabiler zu sein als lexikalische und morphologische.

Für forensische Zwecke erlauben akustische Analysen der *Dialäkt-Äpp*-Aufnahmen beispiellose Erkenntnisse zur arealen Verteilung von Charakteristika des Sprachsignals: Grundfrequenz, Formantfrequenzen, Dauercharakteristika von Segmen-

ten und Sprechgeschwindigkeit können für eine hohe Anzahl Sprecher pro Ortschaft eruiert werden. Die forensisch-phonetische Forschung braucht zur Abklärung darüber, wie (ausser)gewöhnlich die sprachlichen Merkmale eines Täters sowie eines Verdächtigten sind, Populationsdaten, welche auf einer hohen Anzahl von Sprechern mit wenig bis keiner regionalen oder sozialen Variation basieren. Bisher existieren solche Populationsstatistiken nur für gewisse Sprachen (Englisch, Standarddeutsch) und basieren typischerweise auf 50–100 Sprechern (Künzel, Masthoff & Köster, 1995; Jessen, 2007). Eine erste Analyse der *Dialäkt-Äpp*-Aufnahmen von 115 Benutzern aus der Stadt Bern und 205 Benutzern aus der Stadt Zürich zeigt, dass Berndeutsch-Sprecher signifikant langsamer sprechen als Zürichdeutsch-Sprecher: Die in 6 Wörtern pro Sprecher enthaltene Dauerinformation erlaubt es, die beiden Dialekte zu unterscheiden. Das gefäufte perzeptuelle Stereotyp des langsameren Berndeutschen wird also in unseren Produktionsdaten bestätigt (cf. Leemann, Kolly & Dellwo, im Druck; siehe auch Leemann & Siebenhaar, 2010).

Eine Smartphone-Applikation wie die *Dialäkt Äpp* hat das Potenzial, durch *Crowdsourcing* Daten für Forschung zu Morphologie, Syntax, Semantik, Pragmatik, Wahrnehmungsdiakologie, Spracheinstellungen und vielen anderen Fachgebieten zu sammeln. Zum Beispiel könnten Daten für die Forschung zu Spracheinstellungen und perzeptiver Dialekterkennung folgen: Quiz erhoben werden: Dem Benutzer wird ein Dialekt-Quiz präsentiert, in welchem er Dialektaufnahmen den entsprechenden Regionen zuordnen und auf einer affektiven Skala beurteilen soll (cf. Werlen, 1985). Die Hauptsache dabei: Wenn das Interesse des Benutzers durch attraktive Inhalte geweckt werden kann, kommt der Forscher zu seinen Daten. Dem Ermitteln geht immer das Vermitteln voraus, erfolgreiches *Crowdsourcing* geht Hand in Hand mit Wissenschaftskommunikation.

4.0 Fazit und Ausblick

Die *Dialäkt Äpp* ist ein vielversprechendes Werkzeug, um einer hohen Anzahl Benutzer aus der gesamten Deutschschweiz dialektologisches Basiswissen zu vermitteln. Gleichzeitig bietet dieses Werkzeug die Möglichkeit, eine grosse Menge an Daten für die Forschung in verschiedenen Fachgebieten zu sammeln. Weiter bilden die mittels der *Dialäkt Äpp* gesammelten Daten eine einzigartige wissenschaftlich-historische Sammlung zeitgenössischer Deutschschweizer Dialekte.

Verschiedene Linguisten haben uns kontaktiert, um das Framework der *Dialäkt App* auf andere Plattformen zu übertragen und Applikationen zur Erkennung von regionalen Varietäten anderer Sprachen (gesamter deutschsprachiger Raum, amerikanisches Englisch, britisches Englisch, Italienisch) zu kreieren. In einem Nachfolgeprojekt an unserem Labor entwickeln wir zurzeit eine Smartphone-Applikation, die den Dialekt ihrer Benutzer aufgrund automatischer Spracherkennung lokalisiert (*Voice App*, cf. www.voiceapp.ch): Der Benutzer spricht die relevanten Variablen in das Mikrofon des Smartphones, statt sie auf dem Bildschirm anzutippen. Zurzeit verwenden wir die flächendeckenden Variantenwahl-Daten, um 16 zeitgenössische Karten arealer Variation – eine Karte pro Variable – zu erstellen, die anschliessend für eine Untersuchung von Sprach- und Lautwandelphänomenen in Deutschschweizer Dialekten mit ihren Pendanten aus dem SDS verglichen werden können. Zum Beispiel könnte anhand der Daten zur Variablen *Kelle* die Ausbreitung der //Vokalisierung in einem engmaschigen Netz von Ortschaften untersucht werden – eine Analyse, die für den Kanton Bern schon unternommen wurde (Hasler, 2013).

Dank

Wir möchten uns ganz herzlich bei 65 Sponsoren bedanken, welche die *Dialäkt App* via *Crowdfunding* – Finanzierung eines Projekts durch viele Privatpersonen (cf. www.wemakeit.ch) – ermöglicht haben. Merci an Iwar Werlen, der mit seiner umfassenden Kenntnis der schweizerdeutschen Dialektlandschaft bei der Kartenauswahl geholfen hat, an Daniel Wanitsch für das Koordinieren des Software-Teils der App, an Guido Gefter, welcher der *Dialäkt App* seine Stimme geliehen hat, an Fiona Hasler und Carina Eisenblatt für die Digitalisierung der 16 Dialektkarten, an Christoph Landolt, der die App-Benutzer jeden Montag mit einem *Wort der Woche* beglückt, und an Ingrid Hove, die Aufnahmen von zahlreichen Benutzern geprüft hat und auf E-mail-Anfragen der Benutzer antwortet. Weiter danken wir den *Dialäkt-App*-Benutzern für Ihre Bereitschaft, an unserer Forschung teilzunehmen.

Literatur

- BFS = Bundesamt für Statistik (2012). *Amthliches Gemeindeverzeichnis der Schweiz*. <http://www.bfs.admin.ch/> (Stand 13.01.2014).
- Christen, H. (1998). Convergence and divergence in the Swiss German dialects. *Folia Linguistica*, 32, 53–67.
- de Decker, P., & Nycz, J. (2011). For the record: Which digital media can be used for sociophonetic analysis? *University of Pennsylvania Working Papers in Linguistics*, 17, 51–59.
- de Vries, N., Davel, M. H., Badenhorst, J., Basson, W. D., de Wet, F. et al. (2014). A smartphone-based ASR data collection tool for under-resourced languages. *Speech Communication*, 56, 119–131.
- Draxler, C., & Steffen, A. (2005). PhatSessionz: Recording 100 adolescent speakers in schools in Germany. *Proceedings of Interspeech 2005*, 04.–08.09.2005, Lisbon, 1597–1600.
- Duden online (2014). „crowdsourcing“. In: Duden online. <http://www.duden.de/> (Stand 28.02.2014).
- Dufau, S., Duñabeita, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D. et al. (2011). Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. *PLoS ONE*, 6, 1–3.
- Elspass, S., & Möller, R. (2003–2014). Atlas zur deutschen Alltagssprache. <http://www.atlas-alltagssprache.de> (Stand 13.01.2014).
- Faber Acoustical (2009). <http://blog.fabracoustical.com/2009/ios/iphone/iphone-microphone-frequency-response-comparison> (Stand 13.01.2014).
- Glaser, E. (2008). Der Wortschatz des Schweizerdeutschen. http://www.ds.uzh.ch/Forschung/Projekte/Schweizer_Dialekte/index.php (Stand 13.01.2014).
- Glaser, E., Christen, H., Friedli, M., & Renn, M. (2013). *Kleiner Sprachatlas der Deutschen Schweiz*. 5. Aufl. Frauenfeld/Stuttgart/Wien: Huber.
- Hasler, F. (2013). Die Berner Dialektlandschaft im Wandel. Eine vergleichende Untersuchung sprachgeographischer Strukturen. MA-Arbeit, Universität Bern.
- Iwaidja Inyman Team (2012). *Mal Iwaidja*. <https://itunes.apple.com/au/app/mal-iwaidja/id557824618?mt=8> (Stand 13.01.2014).
- Jessen, M. (2007). Forensic reference data on articulation rate in German. *Science and Justice*, 47, 50–67.

- Juska-Bacher, B. (2010). Wortgeografischer Wandel im Schweizerdeutschen. *Sommersprossen, Küchenzwiebel und Schmetterling* 70 Jahre nach dem SDS. *Linguistik online*, 42, 19–42.
- Juska-Bacher, B., Biemann, C., & Quasthoff, U. (2012). Webbasier-te linguistische Forschung: Möglichkeiten und Begrenzungen beim Umgang mit Massendaten. *Linguistik online*, 61, 7–29.
- Katz, J. (2013). Beyond "soda, pop, or coke". Regional dialect variation in the continental US. <http://www4.ncsu.edu/~jakatz2/files/dialectposter.png> (Stand 13.01.2014).
- Künzel, H., Masthoff, H., & Köster, J. (1995). The relation between speech tempo, loudness, and fundamental frequency: an important issue in forensic speaker recognition. *Science and Justice*, 35, 291–295.
- Lameli, A. (2004). *Standard und Substandard. Regionalismen im diachronen Längsschnitt*. Stuttgart: Steiner.
- Landolt, C. (2012–2013). Wortgeschichten auf der Homepage des Idiotikons. <http://wortgeschichten.idiotikon.ch> (Stand 28.02.2014).
- Leemann, A., & Siebenhaar, B. (2010). Statistical modeling of F0 and timing of Swiss German dialects. *Proceedings of Speech Prosody 2010*. Chicago.
- Leemann, A., & Kolly, M.-J. (2013). Dialäkt App. <https://itunes.apple.com/ch/app/dialakt-app/id606559705?mt=8> (Stand 13.01.2014).
- Leemann, A., Kolly, M.-J., & Dellwo, V. (im Druck). Crowdsourcing regional variation in speaking rate through an iOS app. *Proceedings of Speech Prosody 2014*, 20.–23.05.2014, Dublin.
- Leemann, A., Kolly, M.-J., Werlen, I., Britain, D., & Studer-Joho, D. (im Druck). The diffusion of /l/-vocalization in Swiss German. *Language Variation and Change*.
- Mather, P. (2012). The social stratification of /r/ in New York City: Labov's department store study revisited. *Journal of English Linguistics*, 40, 338–365.
- MobiThinking (2013). <http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats/#mobile-app-flops> (Stand 13.01.2014).
- Munro, R., Bethard, S., Kuperman, V., Lai, V. T., Meinick, R. et al. (2010). Crowdsourcing and language studies: the new generation of linguistic data. *Proceedings of the NAACL HLT Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, June 2010, Los Angeles, USA*: 122–130.
- Overeem, A., Robinson, J. C. R., Leijnse, H., Steeneveld, G. J., Horn, B. K. P. et al. (2013). Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters*, 40, 4081–4085.
- Parent, G., & Eskenazi, M. (2011). Speaking to the crowd: looking at past achievements in using crowdsourcing for speech and predicting future challenges. *Proceedings of Interspeech*, 28.–31.08.2011, Florence, Italy: 3037–3040.
- Phonogrammarchiv der Universität Zürich (2000). *Der sprechende Atlas. „Gespräch am Neujahrstag“ in 24 Dialekten*. Frauenfeld: Huber.
- Schlippe, T., Ochs, S., & Schultz, T. (2014). Web-based tools and methods for rapid pronunciation dictionary creation. *Speech Communication*, 56, 101–118.
- SDS = *Sprachatlas der deutschen Schweiz* (1962–2003). Bern (I–VI), Basel: Francke (VII–VIII).
- Siebenhaar, B. (2003). Sprachgeographische Aspekte der Morphologie und Verschriftung in schweizerdeutschen Chats. *Linguistik online*, 15, 125–139.
- Vaux, B., & Golder, S. (2003). *The Harvard Dialect Survey*. Cambridge (MA): Harvard University Linguistics Department.
- Wemakeit (2012). <http://wemakeit.ch/projects/dialaekt-aapp> (Stand 13.01.2014).
- Werlen, I. (1985). Zur Einschätzung von schweizerdeutschen Dialekten. In: I. Werlen (Ed.), *Probleme der schweizerdeutschen Dialektologie*. Freiburg i. Üe.: Universitätsverlag, 195–266.

Schweizerisches Idiotikon

Schweizerdeutsches Wörterbuch

Jahresbericht
2013

